



Identifying environmental signals from population abundance data using multivariate time-series analysis

Masami Fujiwara and Michael S. Mohr

M. Fujiwara (masami@fujiwara.us) and M. S. Mohr, Salmon Population Assessment Fisheries Ecology Division, Southwest Fisheries Science Center, National Marine Fisheries Service, 110 Shaffer Rd., Santa Cruz, CA 95060, USA.

Individual organisms are affected by various natural and anthropogenic environmental factors throughout their life history. This is reflected in the way population abundance fluctuates. Consequently, observed population dynamics are often produced by the superimposition of multiple environmental signals. This complicates the analysis of population time-series. Here, a multivariate time-series method called maximum autocorrelation factor analysis (MAFA) was used to extract underlying signals from multiple population time series data. The extracted signals were compared with environmental variables that were suspected to affect the populations. Finally, a simple multiple regression analysis was applied to the same data set, and the results from the regression analysis were compared with those from MAFA. The extracted signals with MAFA were strongly associated with the environmental variables, suggesting that they represent environmental factors. On the other hand, with the multiple regression analysis, one of the important signals was not identifiable, revealing the shortcoming of the conventional approach. MAFA summarizes data based on their lag-one autocorrelation. This allows the identification of underlying signals with a small effect size on population abundance during the observation. It also uses multiple time series collected in parallel; this enables us to effectively analyze short time series. In this study, annual spawning adult counts of Chinook salmon at various locations within the Klamath Basin, California, were analyzed.

Throughout their life history, individual organisms in a population experience various natural and anthropogenic environmental conditions, and the fluctuations in those conditions are reflected in the fluctuations in population abundance. In addition, observed population dynamics are also affected by observational errors. Consequently, fluctuation in population-abundance time series is produced by the superimposition of multiple signals of different sources. This type of complexity makes it difficult to identify the signals in the population time series. Nevertheless, one of the most important objectives in population ecology is to identify the factors affecting population dynamics (Coulson et al. 2001), and one of the most common data types available in ecology is the time series of abundance.

The superimposition of multiple signals is common in ecological population studies, but analytical methods to handle it are still limited. This limitation results mainly from two statistical problems. First, because population time series are often short, the power of statistical analysis is low. The low power prevents us from fitting complex mechanistic models that include multiple processes. Second, typical regression and related analyses focus on trying to explain the variance of population time series in terms of environmental variables. This does not allow us to identify environmental factors that had small effects on population abundance during a sampling period. However, the

seemingly insignificant environmental factors during a sampling period could become important in the future if dominant environmental factors switch. Therefore, failure to identify the environmental factors that had a small effect-size on population fluctuation will reduce our ability to predict future population abundance.

In the present analysis, a multivariate time-series method called maximum autocorrelation factor analysis (MAFA, Solow 1994) was applied to the multiple population abundance time series. MAFA is one of the approaches that use multiple time series collected in parallel. This alleviates the problem associated with the shortness of time series. MAFA also extracts signals that have high lag-one autocorrelations (i.e. smooth signals, Solow 1994) rather than high variances. This increases a chance to identify the signals that had a small effect size on population abundance. The use of lag-one autocorrelations is especially suitable for ecological population analyses because population time series often have a positive autocorrelation (Pimm and Redfearn 1988, Inchausti and Halley 2001). The autocorrelation can result from environmental variables, which tend to have positive autocorrelations (Vasseur and Yodzis 2004, Ripa and Ives 2007, Fujiwara 2008a), and the life-history strategy of the organism, which can reinforce the positive autocorrelations (Discussion). MAFA takes advantage of these characteristics and extracts environmental

signals from population data (Supplementary material Appendix 1).

In order to demonstrate the utility of MAFA, the technique was applied to the time series of the spawning Pacific salmon abundance (escapement) measured at different locations within the same river basin. Using MAFA, we extracted underlying signals that are common among the locations as well as those specific to particular locations. These signals were then further analyzed by comparing them with various indices that represent environmental conditions thought to affect the dynamics of Pacific salmon. We focused our study on explaining the fluctuation in population abundance of Pacific salmon within the same river basin. Because all of the time series represent the same species and their life history is known, the study is specific about the timing of population processes and environmental events (Supplementary material Appendix 2). This allowed us to gain further insights into the associations between population abundance and environmental variables.

Although MAFA is not new to ecology (Solow 1994, Erizini 2005, Fujiwara 2008b), it has rarely been used compared with other statistical methods. Therefore, our main objective of this study was to reemphasize its advantages in population ecology. In this study, we also analyzed the same data set with a multiple regression analysis, which is one of the most common statistical methods used to associate population time series and environmental variables. Comparing the results from the regression analysis with those from MAFA, we demonstrate the shortcoming of the multiple regression analysis and strengths of MAFA.

Data and methods

Abundance of fall-run Chinook salmon

The analysis is based on the annual adult escapement (age 3, 4 and 5) of fall-run Chinook salmon. As described in Groot and Margolis (1991) and Quinn (2005), this species exhibits an indeterminate semelparous life history. Fall-run Chinook salmon in the Klamath Basin return from the ocean to their natal rivers for spawning between August and November each year. The age of maturity varies between 2 and 5 years old, but most individuals return at age 3 or 4. After spawning, the adults die. Their offspring emerge from the gravel during winter and spring, and after rearing for a short period of time in freshwater, they migrate to the ocean during the summer and fall of the same year and spend one to four years at sea. The fishery-take occurs during their immature stages in the ocean as well as during their spawning upstream migration.

The data were obtained from six locations in the Klamath River and its major tributaries along with one time-series representing the total escapement at some other smaller tributaries over the period 1978 to 2006 (Fig. 1). Although some males mature at age 2, the analysis does not include them. The abundance of spawning adults at each tributary was estimated with various methods including redd (nest) counts, mark-recapture of carcasses, and direct counts using video cameras (KRTAT 2001–2007). These

data are available from the Pacific Fishery Management Council (www.pcouncil.org).

A recent decline in the abundance of this population led to the near-complete closure of Chinook salmon commercial fisheries off the coast of California and Oregon in 2006. However, the cause of the decline in the population abundance is not well understood. Pacific salmon tend to return to their natal streams for spawning. However, depending on the spawning stream, the timing of their upstream migration, egg hatching and downstream migration varies (Quinn 2005). Furthermore, some rivers in the study area are affected by regulated water flow due to dams, but others are not. This is also expected to cause responses to the environmental factors to vary among the locations (e.g. individuals spawning in some locations are more prone to ocean than stream conditions and vice versa).

The fisheries affecting Klamath River fall-run Chinook are annually regulated based on their forecasted current abundance, and the harvest goal is set proportional to the predicted number of available fish in most years (Pacific Fishery Management Council 2009). As expected from the management procedure, the total escapement abundance of fall-run Chinook salmon to the basin is significantly positively correlated with the fishery landing data (unpubl.). Consequently, we assume that the observed short-term fluctuation in the escapement is not caused by the fisheries. However, the short time period for which data are available means that it is important not to extend this assumption when explaining the longer-term fluctuation.

The escapement data are plotted in Fig. 2. Escapements at Trinity River (Fig. 2a), Bogus Creek (Fig. 2e), and the Klamath main-stem (Fig. 2f) currently contribute the most to the total natural area escapement within the Klamath Basin, and the variability in the Trinity River escapement alone constitutes more than 75% of the variance in the total escapement of naturally spawning fall-run Chinook salmon. However, in this analysis, instead of examining the contribution of each tributary to the Basin escapement, we focused on the difference in the signals among these time-series data. Therefore, in the subsequent analyses, the escapement data at each location were standardized by subtracting their mean and dividing by their standard deviation.

Environmental indices

To identify associations between salmon escapements and various environmental conditions that can potentially affect the vital rates (growth, survival, maturation, and reproductive output) of salmon, four types of indices that represent river and ocean conditions were selected based on existing literature on salmon population dynamics. The environmental conditions represented by the indices are river flow rate (NRC 2004, Richter and Kolmes 2005), rate of coastal upwelling (Botsford and Lawrence 2002, Scheuerell and Williams 2005, Wells et al. 2006), hatchery returns, and fishery harvest. These indices are listed in Table 1 and briefly described in Supplementary material Appendix 2. All the environmental variables considered in this analysis have positive lag-one correlation, which provided additional motivation for the use of MAFA in this analysis.

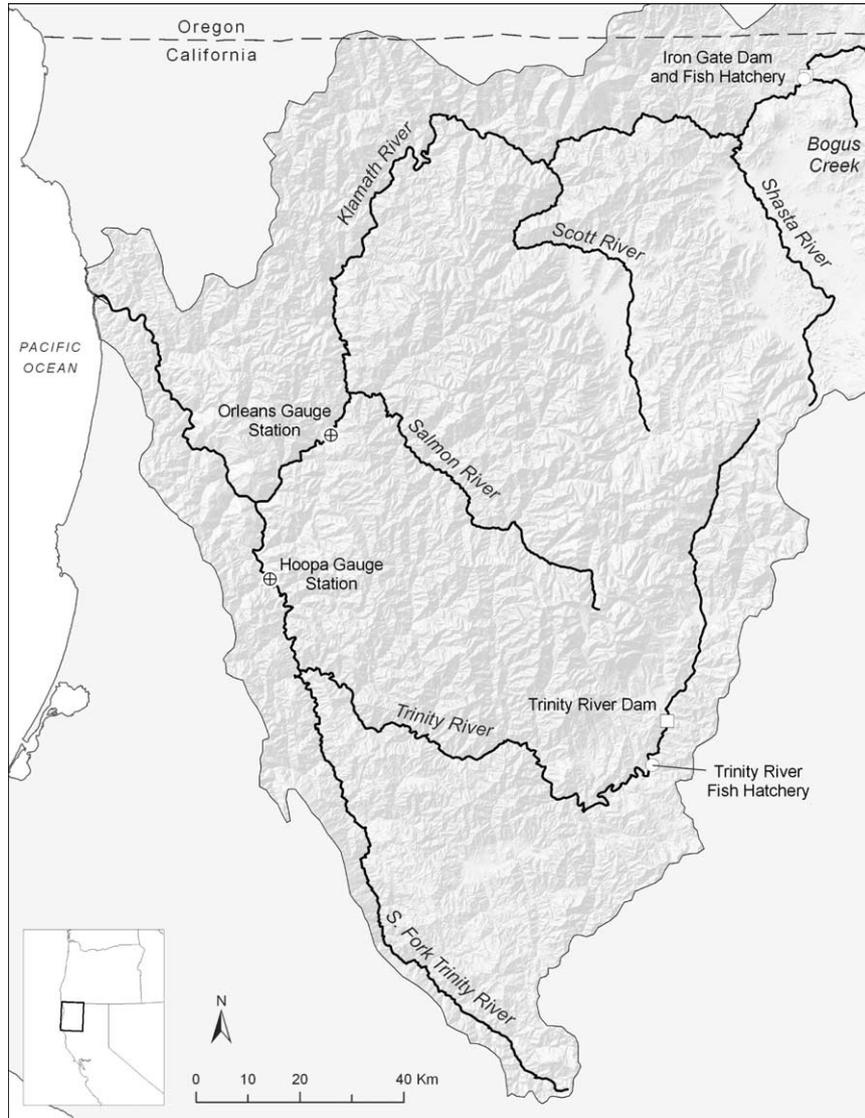


Figure 1. Map of the Klamath Basin, California, USA.

Our preliminary analysis also included Pacific Decadal Oscillation index (Mantua et al. 1997), which is a regional-scale measure of ocean temperature in the North Pacific. However, this index is strongly negatively correlated with the spring coastal upwelling index and the coastal upwelling is expected to have more direct effect on salmon population dynamics (Supplementary material Appendix 2). Therefore, we excluded this index from the final analysis.

Statistical analysis

Maximum autocorrelation analysis

To extract smooth signals, we applied maximum autocorrelation factor analysis (MAFA, Solow 1994) to the escapement time-series data. MAFA is similar to principal component analysis (PCA) in that both methods find weighted linear combinations of the original variables to express new variables that are uncorrelated with each other. However, while PCA finds new variables whose variances are maximized, MAFA finds new variables whose lag-one autocorre-

lations are maximized. Because a smooth signal generally has a high lag-one autocorrelation (Diggle 1990), MAFA identifies new variables that emphasize smooth trends. After standardizing the original escapement data at each location by subtracting its mean and dividing by its standard deviation, the maximum autocorrelation factors (MAFs) were estimated using the algorithm presented by Solow (1994). Let $\tilde{X}_t^{(j)}$ be the standardized j th time-series of the population counts. Then, similarly to PCA, the i th MAF is expressed as a linear combination of the population counts as

$$Y_t^{(i)} = \alpha_{1,i} \tilde{X}_t^{(1)} + \alpha_{2,i} \tilde{X}_t^{(2)} + \dots + \alpha_{m,i} \tilde{X}_t^{(m)} \quad (1)$$

where m is the number of available time series. The coefficients $\alpha_{j,i}$ are estimated by taking the eigenvalue decomposition of the scaled cross lag-one difference matrix (Solow 1994). The sign of the maximum autocorrelation factors were chosen so that they would correlate positively with the total escapement. After obtaining the MAFs, each MAF was standardized by dividing by its standard deviation.

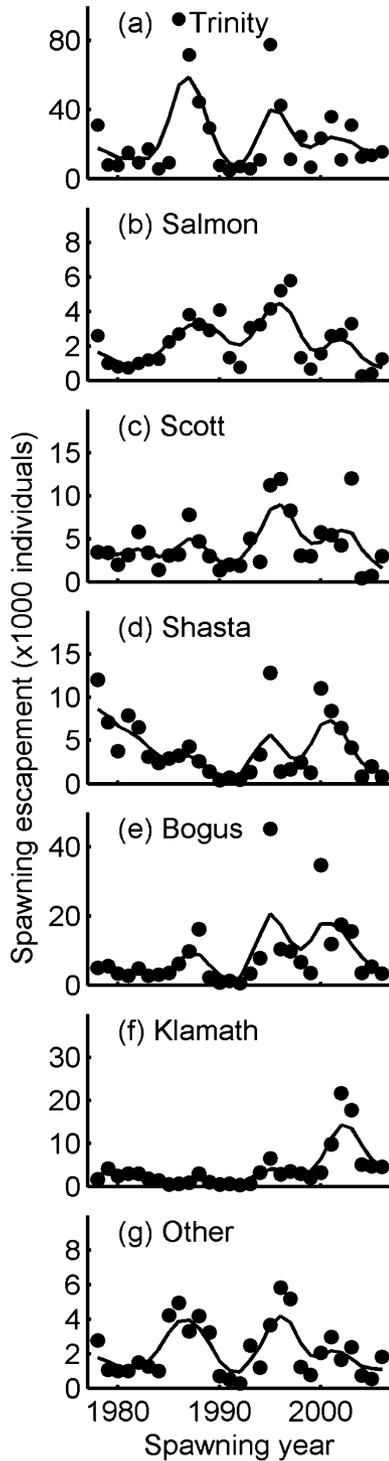


Figure 2. Escapement of fall-run Chinook salmon at (a) Trinity River, (b) Salmon River, (c) Scott River, (d) Shasta River, (e) Bogus Creek, (f) Klamath River main-stem, and (g) other tributaries combined. In each panel, dots are the raw data, and a smooth curve indicates a twice-applied three point moving average.

Let $\tilde{Y}_t^{(i)}$ be the standardized i th MAF. Then, $\tilde{X}_t^{(j)}$ can be expressed in terms of $\tilde{Y}_t^{(i)}$ as

$$\tilde{X}_t^{(j)} = \beta_{1,j}\tilde{Y}_t^{(1)} + \beta_{2,j}\tilde{Y}_t^{(2)} + \dots + \beta_{m,j}\tilde{Y}_t^{(m)} \quad (2)$$

By convention in factor analysis, $\beta_{i,j}$ is called the loading of the j th variable on the i th factor. Because both $\tilde{X}_t^{(j)}$ and $\tilde{Y}_t^{(i)}$ are standardized, $\beta_{i,j}^2$ gives the proportion of variance in the j th variable explained by the i th factor. Finally, the significance of the lag-one autocorrelation of MAF was tested using a one-tailed Bartlett test (Diggle 1990) with a significance level α of 0.05. Only significant MAFs were retained in subsequent analyses as follows:

$$\tilde{X}_t^{(j)} \approx \beta_{1,j}\tilde{Y}_t^{(1)} + \dots + \beta_{k,j}\tilde{Y}_t^{(k)} \quad (3)$$

where k ($k \leq m$) is the number of significant MAFs.

To determine the associations between the MAFs ($\tilde{Y}_t^{(i)}$) and environmental factors, correlations between the MAFs and the indices listed in Table 1 were calculated. Because the MAFs are the weighted linear combinations of the population data, the year associated with them corresponds to the spawning year. On the other hand, the environmental factors affect salmon before spawning, and they have delayed effects on the spawning abundance depending on which life stage is actually affected. Therefore, the years for MAFs and the environmental variables were lagged to match the years of environmental variable and affected spawning abundance (Table 1, Supplementary material Appendix 2).

Multiple regression analysis

The same data set used in the MAFA was also subjected to a multiple regression analysis. In this analysis, each of the seven spawning abundance time series was used as a dependent variable, and environmental variables were used as independent variables. This was repeated for all of the spawning abundance time series. For the escapement at the Trinity River, river flow rate of the Trinity River, hatchery escapement abundance at the Trinity River Hatchery, and all of the coastal upwelling indices were used as independent variables. For the escapements at the other locations, river flow rates of the Klamath River, the hatchery escapement abundance at the Iron Gate Hatchery, and all of the coastal upwelling indices were used. In addition, the total spawning abundance within the basin was regressed against all of the environmental variables. In order to select the independent variables, a forward stepwise regression analysis was used (p -values of 0.05 for inclusion and 0.10 for exclusion).

Results

Maximum autocorrelation factor analysis

Maximum autocorrelation factor analysis revealed three signals (three maximum autocorrelation factors) that have significant lag-one autocorrelation ($p < 0.05$). Hereafter, these factors are denoted as MAF 1, MAF 2 and MAF 3 in the order of decreasing lag-one autocorrelation. In other words, these factors are the weighted linear combination of the spawning escapement data in the order of decreasing smoothness, and these factors are expected to represent underlying environmental fluctuations. The three factors are shown in Fig. 3. The variance of the spawning

Table 1. Environmental variables. Sampling interval is annual. See Supplementary material Appendix 2 for descriptions of the variables.

Variable	Description	Year		
CUM	coastal upwelling index, May	1974–2002	1975–2003	1976–2004
CUJ	coastal upwelling index, June	1974–2002	1975–2003	1976–2004
CUS	coastal upwelling index, September	1977–2005	1978–2006	—
CUO	coastal upwelling index, October	1974–2002	1975–2003	1976–2004
RKF	river flow rate, mean between 1 Oct and 31 Dec, Orleans	1974–2002	1975–2003	—
RKW	river flow rate, mean between 1 Jan and 31 Mar, Orleans	1974–2002	1975–2003	—
RKS	river flow rate, mean between 1 Apr and 30 Jun, Orleans	1974–2002	1975–2003	—
RTF	river flow rate, mean between 1 Oct and 30 Dec, Hoopa	1974–2002	1975–2003	—
RTW	river flow rate, mean between 1 Jan and 31 Mar, Hoopa	1974–2002	1975–2003	—
RTS	river flow rate, mean between 1 Apr and 30 Jun, Hoopa	1974–2002	1975–2003	—
IGH	iron Gate hatchery returns	1978–2006	—	—
TRH	trinity River hatchery returns	1978–2006	—	—
HAV	harvest, all fisheries	1986–2006	—	—

escapement at each location explained by these three factors is: Trinity River (62%), Salmon River (72%), Scott River (41%), Shasta River (61%), Bogus Creek (28%), Klamath River main stem (97%), and other tributaries combined (85%).

Association between MAFs and environmental indices

The three significant MAFs were analyzed further by correlating them with the environmental variables with various time lags. Potential pitfalls of correlation analysis between two serially autocorrelated time series are discussed in Supplementary material Appendix 3. Figure 4 depicts the correlations between the types of the environmental indices and the MAFs. The figure includes those indices with a correlation coefficient magnitude greater than 0.32 with at least one of the two MAFs in each panel (corresponding to a

significance level of $\alpha = 0.1$). The environmental indices are plotted in Supplementary material Appendix 2.

MAF 1 is associated with the spring coastal upwelling indices (Table 2, Fig. 4). The timing of correlated spring coastal upwelling index and MAF 1 (i.e. time lags of 2–4 years) suggested that coastal upwelling during immediately before spawning of the parents of affected cohorts and/or during early life stage of the cohorts themselves in the ocean affects the population.

MAF 2 is associated with fall coastal upwelling. The latter association is apparent from the high correlations between MAF 2 and the fall coastal upwelling indices

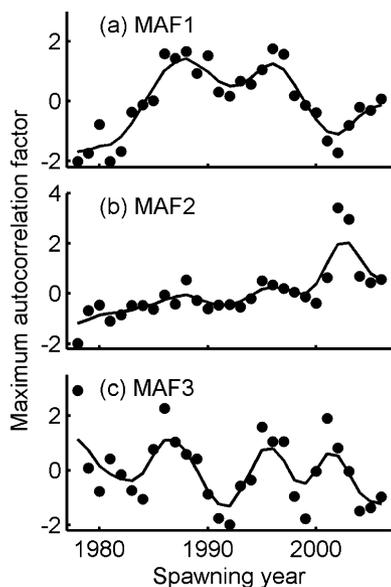


Figure 3. Maximum autocorrelation factors. Dots show resultant linear combinations of original escapement data, and a smooth curve indicates a twice-applied three point moving average.

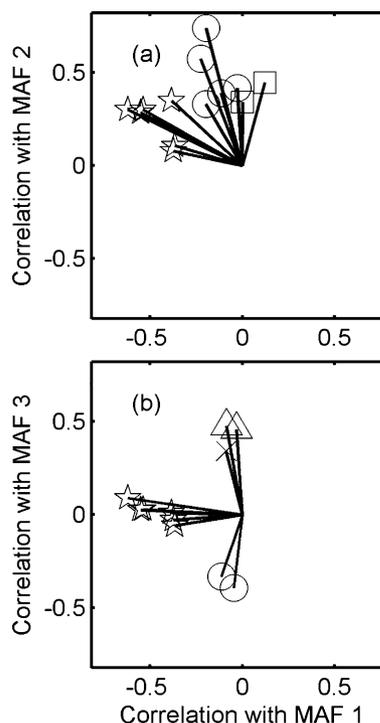


Figure 4. Correlations between MAFs and various environmental indices. \circ : fall coastal upwelling index; \star : spring coastal upwelling index; \times : fall river flow; \triangle : spring river flow; \square : hatchery escapement. Only variables with correlation > 0.32 in magnitude with at least one of the two MAFs in each panel are shown. This level of correlation corresponds approximately to a significance level of 0.10.

Table 2. Environmental variables that are the five most correlated with each of the three maximum autocorrelation factors (MAFs).

	MAF 1	MAF 2	MAF 3
Variable (correlation coefficient)	CUJ, 1974–2002 (–0.62)	CUO, 1977–2005 (0.74)	RTS, 1974–2002 (0.47)
	CUJ, 1975–2003 (–0.55)	CUS, 1975–2003 (0.57)	RKS, 1974–2002 (0.46)
	CUJ, 1976–2004 (–0.54)	TRH (0.44)	CUO, 1975–2003 (–0.39)
	CUJ, 1977–2005 (–0.38)	CUO, 1976–2004 (0.42)	RKW, 1974–2002 (0.34)
	CUM, 1976–2004 (–0.37)	CUS, 1977–2005 (0.39)	CUS, 1976–2004 (–0.33)

(Table 2, Fig. 4). Interestingly, MAF 2 also shows a significant positive correlation with hatchery returns (Fig. 4), suggesting that hatchery fish are also regulated by the same factor.

MAF 3 is mainly associated with the spring river flow rate measures (Fig. 4b). The river flow rate and the escapement data at most locations have negative serial-autocorrelations with lags of 4–6 years, which are evident from the apparent cyclic signals in MAF 3 (Fig. 3c), the river flow data (Supplementary material Appendix 2), and the escapement data (Fig. 2). The latter suggests a strong coherence between escapements and river flow at this frequency. MAF 3 may also be associated with fall coastal upwelling during the early years of ocean residence (Fig. 4b). This suggests that MAF 3 may be a mixture of two environmental variables that were not completely separated by MAFA (Supplementary material Appendix 1).

Grouping of river tributaries

Fig. 5 shows the factor loadings ($\beta_{i,j}$ in Eq. 3) on the three MAFs used to express the natural area spawning escapement time-series ($\hat{Y}_t^{(i)}$; Fig. 1). These factor loadings indicate the relative importance of the MAFs to explain the signals in the escapement time-series. Because the analysis above suggested that MAF 1, MAF 2 and MAF 3 are primarily associated with spring coastal upwelling, fall coastal upwelling, and river flow rate, respectively, we refer to the three factors accordingly in this section. The loadings suggest that the spring coastal upwelling factor (MAF 1) is important for

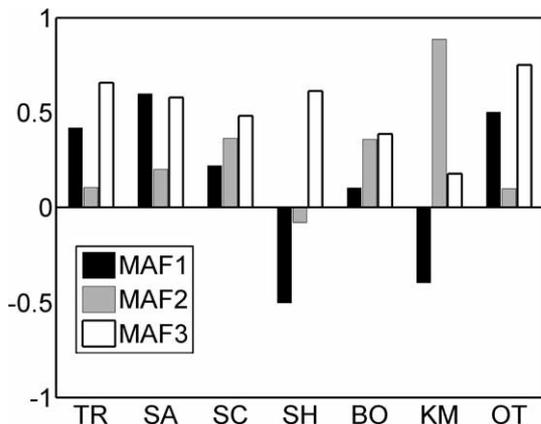


Figure 5. Loadings of the spawning escapements on the three MAFs (TR: Trinity River; SA: Salmon River; SC: Scott River; SH: Shasta River; BO: Bogus Creek; KM: Klamath River main-stem; OT: Other tributaries combined). The loadings were scaled by multiplying by the standard deviation of the corresponding maximum autocorrelation factor.

most locations with the possible exception of Bogus Creek; however, the sign of the association with the escapements differs with the locations. On the other hand, the fall coastal upwelling factor (MAF 2) is the most important factor in explaining the signals at the Klamath River main-stem, and it is also significantly loaded for the Bogus Creek and Scott River escapements. Finally, the river flow rate factor (MAF 3) appears to be important in explaining the observed signals at all locations except the Klamath River main-stem.

The Klamath River and its tributaries are affected by various environmental factors resulting in distinct population dynamics at each location. Based on the association between the escapements and the three estimated factors (Fig. 5), we have categorized the seven tributaries into four groups. The first group includes the Trinity River, Salmon River, Scott River, and the other smaller tributaries, which are primarily correlated with both spring coastal upwelling factor and the river flow rate factor. The second is the Shasta River, which is also affected strongly by the spring coastal upwelling and river flow rate factors, but it has a negative correlation with the spring coastal upwelling factor. The third is the Klamath River main-stem, which is primarily affected by the fall coastal upwelling factor and negatively correlated with the spring coastal upwelling factor. The fourth is Bogus Creek, which exhibits similarity with the hatchery escapements (cf. Fig. 2 and Supplementary material Appendix 3, Fig. 3). However, the variability in escapements at Bogus Creek and the Scott River are explained least by the three maximum autocorrelation factors, which suggests that an additional environmental factor that has small lag-one autocorrelation or sampling variation may be affecting the escapement data at these locations.

Multiple regression analysis

The environmental variables selected by the forward stepwise regression analyses are listed in Table 3. The result suggests that all of the spawning escapement time series except the Salmon River and Scott River escapements are associated with coastal upwelling indices. Furthermore, the Trinity River, Scott River, Shasta River, and Bogus Creek escapements are associated with hatchery escapements. However, differently from the MAFA result, the multiple regression analysis concludes that none of the escapement time series is associated with a river flow rate measure, suggesting that the river flow rate signal is hidden behind other signals.

The total natural-area escapement (all of the escapements combined) is associated with the Trinity River hatchery escapement and a coastal upwelling index. These are two of the variables that are also associated with the

Table 3. Results from multiple regression analysis. Significant independent variables were selected by the forward step-wise linear regression. See text for more detail.

Population data	Environmental variables
Trinity River	CUM, 1976–2004 CUS, 1977–2005 TRH
Salmon River	—
Scott River	IGH
Shasta River	CUJ, 1974–2002 CUJ, 1977–2005 CUS, 1977–2005 CUO, 1976–2004 IGH
Bogus Creek	IGH
Klamath River	CUS, 1975–2003 CUO, 1977–2005
Other tributaries	CUO, 1975–2003
Total basin escapement	CUS, 1977–2005 TRH

escapement at the Trinity River alone. This is consistent with the facts that the variance of the total escapement can be explained mostly by the variance of the Trinity River escapement and that multiple regression analysis selects independent variables that explain the variance of a dependent variable.

Discussion

MAFA as a tool in population analysis

The principal technique used in this study was maximum autocorrelation factor analysis (MAFA), which treats the data as multivariate time-series and extracts smooth signals from them. The technique takes advantage of the fact that a population at different locations may share some common environmental factors but are affected by different amounts. MAFA was particularly effective in identifying the factors affecting the salmon escapements. As anticipated prior to the study, the results suggest that the river condition and coastal upwelling are the important factors affecting the escapement abundance. We also found that the hatchery escapements co-vary with natural-area escapements at some locations. On the other hand, the multiple regression analysis failed to identify the river condition as a potentially important environmental factor. The difference in the results originates from the fundamental difference in the two statistical approaches. MAFA tries to find underlying smooth signals, whereas a multiple regression analysis tries to select independent variables that explain the variance of a dependent variable.

MAFA is particularly suitable for the analysis of population data. First, environmental signals, which affect population processes, tend to have positive lag-one autocorrelation (i.e. they are not purely random). Second, the life-history strategy of organisms can act as a smoothing mechanism. For example, a change in survival probability at one time can affect the population abundance at more than one sampling occasion. Thereby, the environmental signal is smoothed in population data; it is equivalent to applying a low-pass filter. This enhances our ability to extract environmental signals using MAFA. Third, population

time-series are often affected by sampling errors. However, they tend to have a small lag-one autocorrelation and are expected to be included in insignificant MAFs. Finally, long time-series are often not available for population analysis. However, MAFA utilizes the information collected in parallel. This allows the extractions of environmental signals from the data that otherwise cannot be analyzed by other signal processing methods such as Fourier transform and wavelet analyses.

On the other hand, MAFA is not suitable for extracting serially uncorrelated signals. Such signals may become important if organisms are short-lived relative to sampling intervals and the environmental factors affecting the population are serially uncorrelated. In this case, other techniques such as principal component and maximum likelihood factor analyses (Manly 2005) and dynamic factor analysis (Zuur et al. 2003) may be useful.

Pyper and Peterman (1998) caution that correlation analysis between autocorrelated time-series can induce spurious statistical association (type I errors) and suggest the use of a small significance level. In our analysis, both MAFs and environmental signals have a positive lag-one autocorrelation. In fact, the positive autocorrelation is the characteristic of time series that we took advantage of in our analysis. In Supplementary material Appendix 3, we demonstrated that, when time-series consist of multiple signals (i.e. 'noisy time-series'), correlation analysis tends to fail to identify the true associations (increased type II errors). Consequently, we recommend selecting environmental variables carefully based on the prior knowledge of organisms and examining the relative value rather than the magnitude of correlation coefficients (Supplementary material Appendix 3).

In the present study, we took advantage of the spatial structure within a population. Such data may not be available for all populations. However, other types of structure such as stage structure could be used. For example, if the same population is sampled at different developmental stages, the data from different stages could be treated as multivariate time-series data. Such an analysis is advantageous in that one could effectively remove sampling errors, which are expected to have a small lag-one autocorrelation, from the data. Extracted signals could then be examined to investigate the effects of environmental factors on the different stages. This approach would potentially allow us to determine the affected stage.

Klamath fall-run Chinook salmon

The main objective of this paper is to demonstrate the advantages of MAFA in the hope that the results will encourage other studies using MAFA in the future. However, we also believe that the results from our example analysis deserve some discussion. Klamath fall Chinook have experienced both intensive fishing and extensive loss and degradation of suitable freshwater habitat (Myers et al. 1998, NRC 2004). Furthermore, Pacific salmon populations are generally thought to be affected by ocean conditions as well (Mantua et al. 1997, Koslow et al. 2002, Mueter et al. 2002, Logerwell et al. 2003, Scheuerell and Williams 2005, Wells et al. 2006, Zabel et al. 2006).

The fact that there are many potential causes of fluctuations in escapement counts exacerbates the difficulties in understanding salmon population dynamics.

Our analysis revealed three signals in the escapement data. Based on the estimated loadings on the MAFs ($\hat{\beta}_{i,j}$), we categorized the tributaries into four groups. The underlying mechanisms producing these differences are still not clear. However, one potential explanation is the variation in timing of juvenile downstream migration, which can produce variations in the effects of environmental factors. Another possible explanation is the difference in spatial locations. For example, locations adjacent to hatcheries may tend to be affected more by stray hatchery-origin fish (Fig. 1). Regardless of the actual underlying processes, if the differences do result from the differential effects of environmental factors, these subgroups might profitably be considered as sub-units for management purposes, allowing for management at a finer geographic scale.

Cross-correlation analysis between the MAFs and environmental indices showed a number of strong associations. First, MAF 1 is associated with spring coastal upwelling indices. Wells et al. (2006) demonstrated that spring coastal upwelling affects the individual size of returning adult salmon in a tributary of the Smith River, which is adjacent to the Klamath Basin. The results in the present study suggest that spring coastal upwelling may also affect the cohort strength of their offspring. The higher the rate of spring coastal upwelling, the more productive the ocean becomes. This in turn affects the food availability for the fish. Consequently, the spring coastal upwelling strongly influences the energetics of salmon in general. We speculate that spring coastal upwelling during their early ocean life primarily affects the growth and survival of individuals while the spring coastal upwelling during their later ocean life primarily affects their reproductive condition.

Second, MAF 2 is associated with the fall coastal upwelling indices as well as the escapements of hatchery fish. Food availability in the fall may be an important factor affecting the winter survival of salmon in the ocean. Why this factor (but not the other factors) is associated with hatchery escapement is less clear. Further studies to explain or falsify the association are clearly needed.

Finally, MAF 3 is associated primarily with the river flow indices. MAF 3 and many of the river flow rate measures have negative serial-autocorrelations with lags of 4–6 years (i.e. a cyclic signal). Interestingly, similar autocorrelations are observed in the river flow rates at the Smith River and Redwood Creek, whose watersheds are adjacent to the Klamath Basin (unpubl.). Because these two streams are free-flowing (unaffected by dams), this similarity indicates that the autocorrelations are naturally occurring phenomena. Crozier and Zabel (2006) recently presented evidence suggesting that river width and temperature are two important factors determining the survival of juvenile Chinook salmon in the Salmon River Basin, Idaho. It is plausible that the survival of juveniles rearing within the Klamath Basin is also affected by these river conditions.

Salmon population dynamics are often explained by Ricker-type models, which exhibit over-compensatory density dependence. The cyclic signal observed in the

escapement data is approximately the generation time of Chinook salmon. Thus, another possible explanation for such a signal is the over-compensatory density dependence process. However, the result from the present study suggests that it is likely to be caused by a similar signal in the environmental condition. This also suggests the difficulty of identifying density dependence from observed population dynamics.

The multiple regression analysis, on the other hand, did not suggest the importance of river conditions on the salmon population dynamics. This is simply because the fluctuation in the river flow rate did not contribute significantly to the variance of the escapement abundance time series during the sampling period. However, this does not imply that the river conditions do not affect the population dynamics. In fact, we were able to isolate the river flow rate trend from the same data using MAFA. Therefore, if we ignore the effect of river conditions in the future population management, we may encounter surprised changes in the population abundance when the river conditions are changed drastically. This highlights the danger of examining the variance of data alone.

Caveats in interpretations

In the analysis presented above, we selected several indices to represent environmental conditions that have been previously proposed to affect salmon populations. The significant statistical associations support some of the previously proposed processes; however, while this is an important first step toward understanding the underlying mechanisms, we still do not understand the details of these processes.

Although our analysis indicates the importance of river flow in terms of salmon abundance, in representing this particular environmental condition, we used only the mean river flow rates for portions of the year at selected locations. It is likely that a more important measure, which has a more direct effect on the vital rates of salmon, involves other aspects of the river flow characteristics, timing, and/or location (Hilborn et al. 2003). Our approach was based on the idea that the degree of correlation between such a measure and our environmental indices would be sufficiently strong within a broad category (e.g. river flow) that we would be able to detect the association between salmon abundance and the environmental condition. The same logic was used in determining the other environmental indices.

Similarly, we did not include other potentially important measures of environmental conditions such as oxygen content, nutrient concentration, amount of vegetation, water temperature, and rainfall, or the timing of these events, which may show strong correlation with some of our measures. These factors may affect the vital rates of salmon more directly than some of the factors assessed in the present study. As a result, the findings presented in this paper should be considered carefully to avoid over-interpretation.

The caveats described above, however, are the limitations of statistical analysis in general. To become confident about a cause-and-effect relationship, we have to obtain multiple pieces of evidence of different kinds that all support the relationship or we have to obtain multiple pieces of

evidence of the same kind under a wide range of conditions (Cox and Wermuth 2004). Therefore, despite the limitations, we believe the results from a statistical analysis like the one presented in this paper along with our prior knowledge of biological processes provide important pieces of evidence toward understanding the true processes.

Acknowledgements – We thank Peter Adams for suggesting this topic of research, and for constructive comments on the work. We also thank Stan Boutin, Tim Coulson, Timothy Keitt, Steve Lindley and Andrew Solow for their constructive comments and Thomas Pearson and Tommy Williams for their assistance. A special thanks is extended to Klamath River Technical Advisory Team members, past and present, for their comprehensive estimation and documentation of Klamath River fall-run Chinook salmon escapements and fishery harvest, without which this research would not have been possible; in particular, Allen Grover, Melodie Palmer-Zwhalen, George Kautsky, Wade Sinnen, Dave Hillemeier, Jerry Barnes and Desma Williams.

References

- Botsford, L. W. and Lawrence, C. A. 2002. Patterns of co-variability among California Current chinook salmon, coho salmon, Dungeness crab, and physical oceanographic conditions. – *Progr. Oceanogr.* 53: 283–305.
- Coulson, T. et al. 2001. Age, sex, density, winter weather and population crashes in Soay sheep. – *Science* 292: 1528–1531.
- Cox, D. R. and Wermuth, N. 2004. Causality: a statistical view. – *Int. Stat. Rev.* 72: 285–305.
- Crozier, L. and Zabel, R. W. 2006. Climate impacts at multiple scales: evidence for differential population responses in juvenile Chinook salmon. – *J. Anim. Ecol.* 75: 1100–1109.
- Diggle, P. J. 1990. Time series: a biostatistical introduction. – Oxford Univ. Press.
- Erizini, K. 2005. Trends in NE Atlantic landings (south Portugal): indentifying the relative importance of fisheries and environmental variables. – *Fish. Oceanogr.* 14: 195–209.
- Fujiwara, M. 2008a. Effects of an autocorrelated stochastic environment and fisheries on the age at maturity of Chinook salmon. – *Theor. Ecol.* 1: 89–101.
- Fujiwara, M. 2008b. Identifying interactions among salmon populations from observed dynamics. – *Ecology* 89: 4–11.
- Groot, C. and Margolis, L. 1991. Pacific salmon life histories. – UBC Press.
- Hilborn, R. et al. 2003. Biocomplexity and fisheries sustainability. – *Proc. Natl Acad. Sci. USA* 100: 6564–6568.
- Inchausti, P. and Halley, J. 2001. Investigating long-term ecological variability using the Global Population Dynamics Database. – *Science* 293: 655–657.
- Koslow, J. A. et al. 2002. Climate variability and marine survival of coho salmon (*Oncorhynchus kisutch*) in the Oregon production area. – *Fish. Oceanogr.* 11: 65–77.
- KRTAT-Klamath River Technical Advisory Team 2001–2007. Klamath River fall Chinook age-specific escapement, river harvest, and run size estimates. Available from US Fish and Wildlife Service, 1829 South Oregon Street, Yreka, CA 96097, USA.
- Logerwell, E. A. et al. 2003. Tracking environmental processes in the coastal zone for understanding and predicting Oregon coho (*Oncorhynchus kisutch*) marine survival. – *Fish. Oceanogr.* 12: 554–568.
- Manly, B. F. J. 2005. Multivariate statistical methods – a primer (3rd ed.). – Chapman and Hall/CRC.
- Mantua, N. J. et al. 1997. A Pacific interdecadal climate oscillation with impacts on salmon production. – *Bull. Am. Meteorol. Soc.* 78: 1069–1079.
- Mueter, F. J. et al. 2002. Opposite effects of ocean temperature on survival of 120 stocks of Pacific salmon (*Oncorhynchus* spp.) in northern and southern areas. – *Can. J. Fish. Aquat. Sci.* 59: 456–463.
- Myers, J. M. et al. 1998. Status review of Chinook salmon from Washington, Idaho, Oregon, and California. NOAA Technical Memorandum. NMFS-NWFSC-35. – US Dept of Commerce.
- NRC- National Research Council of the National Academies 2004. Endangered and threatened fishers in the Klamath River basin: causes of decline and strategies for recovery. Natl Res. Council Natl Acad. – The National Academies Press.
- Pacific Fishery Management Council 2009. Review of 2008 Ocean Salmon Fisheries. (Document prepared for the Council and its advisory entities.) – Pacific Fishery Management Council, Portland, Oregon, USA.
- Pimm, S. L. and Redfearn, A. 1988. The variability of population densities. – *Nature* 334: 613–614.
- Pyper, B. J. and Peterman, R. M. 1998. Comparison of methods to account for autocorrelation in correlation analyses of fish data. – *Can. J. Fish. Aquat. Sci.* 55: 2127–2140.
- Quinn, T. P. 2005. The behavior and ecology of Pacific salmon and trout. – Univ. of Washington Press.
- Richter, A. and Kolmes, S. A. 2005. Maximum temperature limits for chinook, coho, and chum salmon, and steelhead trout in the Pacific Northwest. – *Rev. Fish. Sci.* 13: 23–49.
- Ripa, J. and Ives, A. R. 2007. Interaction assessments in correlated and autocorrelated environments. – In: Vasseur, D. A. and McCann, K. S. (eds), The impact of environmental variability on ecological systems. Springer, pp. 111–130.
- Scheuerell, M. D. and Williams, J. G. 2005. Forecasting climate-induced changes in the survival of Snake River spring/summer Chinook salmon (*Oncorhynchus tshawytscha*). – *Fish. Oceanogr.* 14: 448–457.
- Solow, A. R. 1994. Detecting change in the composition of a multispecies community. – *Biometrics* 50: 556–565.
- Vasseur, D. A. and Yodzis, P. 2004. The color of environmental noise. – *Ecology* 85: 1146–1152.
- Wells, B. K. et al. 2006. Covariation between the average lengths of mature coho (*Oncorhynchus kisutch*) and Chinook salmon (*O. tshawytscha*) and the ocean environment. – *Fish. Oceanogr.* 15: 67–79.
- Zabel, R. W. et al. 2006. The interplay between climate variability and density dependence in the population viability of Chinook salmon. – *Conserv. Biol.* 20: 190–200.
- Zuur, A. F. et al. 2003. Estimating common trends in multivariate time series using dynamic factor analysis. – *Environmetrics* 14: 665–685.

Supplementary material (available online as Appendix O17570 at www.oikos.ekol.lu.se/appendix). Appendix 1–3.